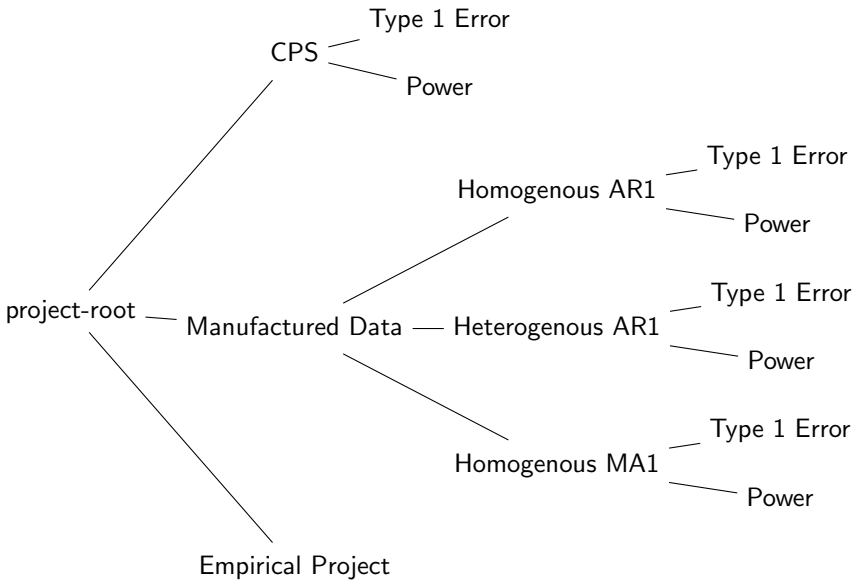
README

1 Project folder organisation



Along with these files the project root also contains the README and the hansen.ado files.

Before running the FGLS and BC-FGLS files, the hansen.ado file (provided at the project root) has to be run in Stata. This is required to initiate the hansen program in stata.

All of our analysis, are perfectly reproducible, as we have conditioned for all avenues of randomness in the code. But the exact answers may vary a little (to the second decimal point) because of the different versions of Python and Stata, and different versions of the modules (such as Pandas, Numpy) used in the Python codes.

2 CPS DATA :

1. Each of the folders of Type1_Error and Power contain the following files:

   OLS,

   CRSE_N(0,1),

   CRSE_t(G-1),

   Residual_Aggregation,

   Bootstrap,

   FGLS,

   BC-FGLS.

2. Out of all the files FGLS and BC-FGLS are done in Stata while all others are done in Python.

3. The code is created in such a way that all the files(both under Power and Type1_Error) can take in different numbers of states to suit our analysis.

4. In the CPS DATA, the maximum number of states possible is 50.

2.1 How to use:

This instruction is valid for all the Python and Stata files under both the Type 1 and Power folder of CPS DATA.

1. In the first cell of each notebook, the file path for the CPS_data.csv has to be introduced in the variable file_path. The file path has to be given in a specific format such as r"file_path". In all of the files, one of our file paths for the data is provided which needs to be changed accordingly.

2. In the do files of Stata, the file path has to be given in conjuction with the 'use' command. We have mentioned in the code, where the file path needs to be provided. One of our file paths is provided in the codes we provided.

3. The variable num_of_states can be found in the second cell of each notebook for all the Python files. It can be changed according to the desired number of states.

4. In the do files of stata, the variables num_of_states have to be changed to the desired number of states. The variable can be found in the second line of the code after the command 'clear'.

3 MANUFACTURED DATA :

1. Each of the folders of Type1_Error and Power contain the following files:

   OLS,

   CRSE_N(0,1),

   CRSE_t(G-1),

   Residual_Aggregation,

   Bootstrap,

   FGLS,

   BC-FGLS.

2. Out of all the files FGLS and BC-FGLS are done in Stata while all others are done in Python.

3. The code is created in such a way that all the files(both under Power and Type1_Error) can take in different numbers of states to suit our analysis.

4. In homogenous AR1, number of states is not a constraint, as the data will be generated for the desired number of states in the first place.

5. The folders under Heterogenous AR1 and Homogenous MA 1 do not support different numbers of states, as we have carried out our analysis taking always 20 states for these 2 data generation methods.

3.1 How to use:

This instruction is valid for all the Python and Stata files under both the Type 1 and Power folder of MANUFACTURED DATA.

1. There is no provision for the file_path variable in the files of MANUFACTURED DATA, as the data is generated not imported.

2. The data generation process supports the generation of data for any number of states and any number of time periods through the variables N and T respectively. For all our analysis we have used T = 20, but the variable N can be changed to the desired number of states.

3. For the Python files, The variables N and T can be found again in the first cell of notebooks.

4. In the do files of Stata, The variables N and T can be found from the second line of the code just after the command 'clear'

5. For homogenous AR1, rho can be changed to different values according to the requirements of the analysis. It can also be found in the first cell of each notebook under Homogenous AR1.

6. For heterogenous AR1 folder, there is no scope of rho as rho is randomly picked from U (0.2, 0,9) for each state and for homogenous MA(1) all the analysis has been carried out for theta = 0.5

4 EMPIRICAL PROJECT

This folder has 4 files :

1. Data_for_Clustering.xlsx : This is an excel sheet containing the data of the socioeconomic variables as imported from World Bank (World Development Indicators)

2. Clustering.ipynb : This contains all the codes regarding the data preperation of the socioeconomic variables for the clustering and the process of clustering.

    - In the first cell the file path of Data_for_Clustering.xlsx has to be given to run the notebook.

3. Final_Regression_Data.dta : It contains the final data of the Agri Imports of each country for each year (sourced from FAOSTAT), including the Treatment variable indicator and Cluster assignment. This is the data on which the final regression will be run.

4. Final_Regression_Agri_Imoorts.do : This is the do file containing the codes for the final regression.

    - The file path for Final_Regression_Data has to be provided.
    - The Cluster for which the regression has to be run, needs to be specified, in the relevant area as mentioned in the code.